July 1970

Dr. J. Donohue
University of Pennsylvania
Department of Chemistry
Philadelphia, Pennsylvania 19104
U.S.A.

Dear Jerry

After reading your letter of 9 June I think we had better go back and start at the beginning.

First a remark about "effectively centric" reflections. These are those for which there are only two alternative choices for the phase, rather than a continuous choice. It is normally arranged so that this choice is either 0 or π, but since phase is partly a matter of convention (since the origin is strictly arbitrary, although convenience may suggest the easiest choice) this is not a strict requirement. For a structure like DNA, a reflection does not necessarily have to have $B = 0$ to be "effectively centric".

However, the important question, to which I do not yet know the exact answer, is what fraction of the intensity of a DNA diffraction pattern comes from effectively centric reflection? I have asked the people at King's College to look into this. When I know the answer I will write to you again.

Next I must take up a point from your earlier letter. We would both agree that whether a structure is centric or not it is possible, under certain circumstances, to arrive at a false solution. However, it is the general opinion of crystallographers that it is significantly easier to make a mistake in an acentric structure than in a similar centric one. You appear to deny this, but I cannot believe you really mean it. To make matters precise, do I have your permission to quote your opinion as follows? "I understand from Donohue (personal communication) that he considers it just as easy to arrive at a false solution for a centric structure as for a similar acentric one".

Next I should like to make some general remarks about solving crystal structures. In the first place, in spite of the excellence of the x-ray pictures taken at King's, the resolution is somewhat limited compared to the single crystals of most small organic molecules. Thus, even if the phases were known rather accurately it is unlikely that atoms would be resolved enough so that the structure could be deduced without knowing the chemistry of the material. I think you would in any case argue that it is fair to accept the

continued ...

chemistry as given. The basic reason why this is often acceptable in crystallography is that there is often no reasonable doubt about the chemistry. So much so that if a first structure contradicts what is firmly established most people take another look at the structure. In fact, you yourself have done much useful work in spotting incorrect structures, using just this technique.

You will no doubt want to point out to me that this has dangers. That unless the chemical evidence is overwhelming it can happen that mistakes will creep in through assuming chemical facts which were actually incorrect. Moreover, you could say that the fact that a structure does not fit our preconceptions is a hint that it may be wrong, not a proof. Nevertheless, it is I think legitimate to take certain "chemical" facts as given when trying to solve a structure, and I hope you will agree that this, in principle, is acceptable. If so, it remains to set out what facts can be accepted in the case of DNA.

Now the general chemical formula (as opposed to the detailed base sequence) is surely known beyond reasonable doubt, not only from many chemical studies, but also for crystal structures of small, related, molecules. Thus, the "topology" of the molecule is known - i.e. what is connected to what by chemical bonds (as opposed to hydrogen bonds, etc.)?

Next the length of all the chemical bonds is known to sufficient accuracy. That is, if one were in fact incorrect it would be by such a small amount that it is highly unlikely to affect the problem. The bond angles are also known, but not to such a high degree of precision. However, I doubt if anyone would accept without special justification a "tetrahedral" angle of 135^{o}, or 75^{o}.

About the dihedral angles one must be more careful. Although the bases may not be strictly quite planar the deviation from planarity is likely to be so small that it will not cause any trouble. On the other hand, although both you and I would be unhappy if the C_4 - C_5 bonds were in the eclipsed configuration, I think it would be wiser <u>not</u> to forbid this in the first place and to let the dihedral angle have any value. The same would be said about the planarity or otherwise of the ribose ring. In short, I should be surprised if we could not sit down and work out a set of acceptable rules for the chemical bonds. Moreover, these rules would probably approximate very closely to those actually used at King's.

It follows that the totality of "acceptable" configurations is defined by rather few parameters. These are, in fact, the dihedral angles and the position and orientation of the "monomer" in the unit cell. It is difficult to estimate the number of "effective" parameters because of the constraints (e.g. the monomers must join head to tail, the ribose ring must be closed, atoms should not penetrate, etc.) but it is quite small.

The chemical structure is not the only thing that can now be assumed as being beyond reasonable doubt. There are three other facts which originally

Dr. J. Donohue

could not be assumed, but now have to be considered. They are:

(1) The bases "pair" - the exact way need not be specified - such that A pairs with T, and G with C.

(2) There are two chains in the structure.

(3) The (pairing) chains run in opposite directions.

I shall argue that (1) and (3) are established beyond reasonable doubt, and that (2) is probable though not perhaps quite certain.

The original evidence for "pairing" of the bases came from the base composition of DNA. This evidence still stands, supplemented by the additional evidence that single-stranded DNA does not usually show one-to-one ratios. However, the overwhelming evidence comes from the synthesis of synthetic DNA, mainly by Khorana. The original evidence was from repeating sequences, but more recently, in synthesising the DNA for a tRNA gene he has constructed many non-repeating sequences. These firmly establish that for two chains to go together the pairing rules must be obeyed, and also that the two sequences must run in opposite directions - he has even done a special experiment to prove that if the complementary sequence runs in the same direction the chains do not pair. These experiments do not, by themselves, say what the mechanism of pairing is, nor do they rule out the possibility of other base pairs in small amounts, but since the x-ray data looks at the average structure these can safely be ignored for the time being. There is, of course, lots of other evidence which suggests base-pairing for RNA, such as the relationship between codon and anticodon, and the paired regions in tRNA (the latter is basically a statistical argument). Both these lines of evidence, incidentally, suggest that G - U pairs can occur. However, I am ignoring all this data as too indirect.

The question of the number of chains is not really in serious doubt, but because of Cavalieri there has been some controversy about it. What is not in doubt is the mass per unit length. This can, of course, be obtained direct from the crystal data. It can also be deduced from such techniques as the sedimentation velocity in solution, or the length of a piece of DNA of known molecular weight as seen in the electron microscope. All these methods agree that the value is near 200 daltons/A, or two residues per 3A, approximately. However, it is just possible that the structure has four chains, each with a residue every 6A or so. The best evidence for two chains is from circular DNA molecules which form super coils, and which relax to a simple circular form after a single break. There is a lot of other indirect evidence, but I am not sufficiently up-to-date in the subject. If you insisted that four-chain models should not be excluded I would have to agree with you, while saying that I think them very unlikely.

Now we must come to the screw axis and the symmetry. Do you have any serious doubts about the helical parameters (except for RNA, where there is an uncertainty)? The argument depends on the absences in the pattern, and is therefore statistical, but then, curiously enough, so are all arguments for the deduction of symmetry from the x-ray data. The only question is, is the evidence strong enough? It seems to me it is, but perhaps you have other views.

I would not wish to argue that the symmetry of the unit cell can be deduced from intensity statistics, since this is rather technical. I would take the view that since other evidence shows the chains run in opposite directions, and since the shape of the unit cell and the intensity statistics are at least compatible with this, it would be highly surprising if the molecule did not have (pseudo) dyads perpendicular to the fibre axis, and rather unlikely that the cell would not use them. Obviously models having parallel chains, with molecules statistically up and down need not now be considered.

What then, is required from the x-ray data, since so much of the general properties of the structure have been established by other methods? What the x-ray data is required to do is

(1) show that it is compatible with a possible model
(2) establish the stereochemical nature of the base-pairing.

Now, as you know, base-pairing has been observed in single crystals of small molecules. These show the G - C pair to be as predicted, but the A - T pair usually in a different form, using the N_7 of adosine. Thus, the exact pairing of the bases is a problem, and it would be most useful if we could say that the x-ray data made it highly probable.

We now come to the question set out in your letter. "Does model building followed by electron density calculations furnish proof of a structure?" I think this question is, at the moment, in too strong a form. We would agree that if the effective parameters were few, the experimental data very extensive, and the result completely acceptable stereochemically, then the structure was probably correct. You have had a lot of experience with small crystals, but practically none with such awkward polymers as DNA. Thus, it is very difficult to arrive at criteria which we can agree are acceptable in this case. That is why any alternative model, preferably produced by someone other than the original worker, is preferable. (See the history of the structure of collagen if you want an amusing example!) However, we have a more limited problem immediately before us. Can the x-ray data for DNA eliminate a model, and in particular one with a differen~~ce~~t, specified base-pairing?

The King's people claim they can eliminate it. I claim that your counter-examples are not relevant, because you over-simplify the problem by considering only base-pairs, and not the complete structure, and because you consider acentric examples, whereas the DNA structure is, to some extent, effectively centric. To what extent it is effectively centric and how much this matters remains to be determined.

However, I have established to my own satisfaction that you do not yet have an adequate theoretical grasp of helical diffraction theory, that your arguments from hypothetical models of base-pairs are theoretically not to be trusted without further justification, that the theoretical position is complicated and indeed likely to lead to "long acrimonious arguments", whereas the production of an acceptable alternative model by you would settle the argument in your favour. I still cannot see any other way for you to prove your point. What about it? After all, it would be very useful experience for you to have a shot at it. I am sure I could persuade the King's group to make the experimental data available if you have not got it already.